# Structural basis for regulation of rhizobial nodulation and symbiosis gene expression by the regulatory protein NolR

Soon Goo Lee[a], Hari B. Krishnan[b], and Joseph M. Jez[a,1]

[a]Department of Biology, Washington University, St. Louis, MO 63130; and [b]Plant Genetics Research Unit, United States Department of Agriculture-Agricultural Research Service, University of Missouri, Columbia, MO 65211

Edited* by Eva Kondorosi, Hungarian Academy of Sciences, Biological Research Centre, Szeged, Hungary, and approved March 19, 2014 (received for review February 10, 2014)

**The symbiosis between rhizobial microbes and host plants involves the coordinated expression of multiple genes, which leads to nodule formation and nitrogen fixation. As part of the transcriptional machinery for nodulation and symbiosis across a range of Rhizobium, NolR serves as a global regulatory protein. Here, we present the X-ray crystal structures of NolR in the unliganded form and complexed with two different 22-base pair (bp) double-stranded operator sequences (oligos AT and AA). Structural and biochemical analysis of NolR reveals protein–DNA interactions with an asymmetric operator site and defines a mechanism for conformational switching of a key residue (Gln56) to accommodate variation in target DNA sequences from diverse rhizobial genes for nodulation and symbiosis. This conformational switching alters the energetic contributions to DNA binding without changes in affinity for the target sequence. Two possible models for the role of NolR in the regulation of different nodulation and symbiosis genes are proposed. To our knowledge, these studies provide the first structural insight on the regulation of genes involved in the agriculturally and ecologically important symbiosis of microbes and plants that leads to nodule formation and nitrogen fixation.**

transcription factor | protein structure

The symbiosis between rhizobial bacteria from the *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Azorhizobium*, and *Bradyrhizobium* genera and leguminous plants leads to the formation of root nodules (1, 2). These plant organs are specialized for nitrogen fixation and assimilation and are of major ecological and agricultural importance. For example, nitrogen-fixing nodules account for one quarter of total nitrogen fixed globally each year (3). The development of nitrogen-fixing nodules by rhizobia involves a variety of interactions between the plant and microbe; however, at the center of this process are a set of *nod* (nodulation) genes required for the synthesis of oligosaccharide-nodulation factors, for determining host-plant specificity, and for optimizing the efficiency of symbiosis (4–6). Successful interaction between the rhizobium and host plant requires expression of both positive and negative transcriptional control of genes related to nodulation and symbiosis (7, 8).

Expression of *nod* genes in rhizobium is regulated by flavonoids released from the host plant in conjunction with the positive activator NodD, which is a member of the LysR-type transcriptional regulator (LTTR) family of proteins (9, 10). Extensive analyses reveal that, in response to small molecules produced by the plant, the rhizobial NodD protein binds to a *cis*-acting element—the *nod* box—located upstream of genes required for nodulation (11–16). In addition to the positive control provided by NodD, negative regulation of the *nod* regulon in *Rhizobium meliloti* and other rhizobia by NolR occurs (17, 18).

In rhizobia, NolR modulates expression of the NodD activator protein, the core *nod* genes, and multiple genes involved in symbiosis (17, 19–24). Based on amino acid sequence homology, NolR was proposed to be a helix-turn-helix family member that binds to a nonpalindromic consensus motif —(A/T)TTAG-$N_9$-A(T/A) (17). NolR is well-conserved across multiple *Sinorhizobium* and *Rhizobium* species (23). Originally, NolR was identified as a negative regulator of nodulation that bound to overlapping transcription initiation sites in the *nodD1* and *nodA* promoters and at the *nodD2* promoter (17). Differential regulation of *nod* genes by NolR suggested that only the genes related to the synthesis of the core Nod factor structure were controlled by this transcription factor (22). Interaction of NolR with the target DNA site led to reduced levels of Nod factors (22). Control of Nod factor production by NolR may also aid in optimization of nodulation specificity, as NolR binding sequences were later found in the promoter regions of *nodABC*, *nodD1*, *ttsI-nodD2*, *nolR*, *hesB*, and *nodZ* (23). Transcript levels of *nolR* are high in free-living rhizobia and in the bacteroid but are down-regulated by luteolin, a *nod* gene inducer (24). Subsequent studies implicated NolR as a global regulatory factor that responds to environmental factors to fine-tune a range of symbiotic signals, not just the genes required for nodulation, and that the absence or presence of NolR affects symbiotic interactions with host plants (19, 20, 23, 24). For example, NolR represses expression of the type III secretion system *ttsI* gene, which is required for secretion of nodulation outer proteins (nops) that are beneficial for symbiosis of *Sinorhizobium fredii* with some soybean cultivars (24). The molecular basis for NolR recognition of nonpalindromic DNA target sites and subsequent control of nodulation and symbiosis genes is unclear.

To understand how NolR functions as a global regulator of nodulation, we used a combination of X-ray crystallography,

## Significance

**Nitrogen nodules formed by the symbiosis of rhizobial microbes and legume roots are essential for fixation of nitrogen in the environment. As part of the symbiosis that leads to nodule formation, a series of changes in gene expression of the Rhizobium must occur. The protein NolR is a global regulator of rhizobial genes for symbiosis and nodulation. Here, we describe the three-dimensional structure of this transcription factor in unliganded and DNA bound forms. These structures show how NolR recognizes asymmetric DNA binding sites and reveal a previously unknown mechanism for conformational switching that alters the energetics of interaction to accommodate variable DNA sequences. Two models for the role of NolR in the regulation of nodulation and symbiosis genes are also proposed.**

Author contributions: S.G.L., H.B.K., and J.M.J. designed research; S.G.L. performed research; S.G.L. and J.M.J. analyzed data; and S.G.L., H.B.K., and J.M.J. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org [PDB ID codes 4OMY (SeMet NolR•oligo AT DNA), 4OMZ (unliganded NolR), and 4ON0 (NolR•oligo AA DNA)].

[1]To whom correspondence should be addressed. E-mail: jjez@wustl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1402243111/-/DCSupplemental.

thermodynamic analysis of protein–DNA interactions, and site-directed mutagenesis. The structures of NolR in the unbound form and in complex with two different 22-base pair (bp) double-stranded DNA fragments (oligos AT and AA) reveal a homodimeric protein adopting a winged helix-turn-helix fold. These structures suggest a previously unknown mechanism for conformational switching of a key glutamine side-chain to accommodate DNA-sequence variation in nonpalindromic operator sites without a loss of interaction affinity but with altered thermodynamic contributions to binding. Models for the role of NolR in the regulation of nodulation and symbiosis genes are proposed.

## Results

**Overall Structure of NolR.** NolR from *S. fredii* USDA191 was expressed as an N-terminally His-tagged protein in *Escherichia coli* and purified by nickel-affinity and size-exclusion chromatography. The His-tag was removed by thrombin digestion for crystallization. NolR migrated as a dimeric 26-kDa species (monomer $M_r \sim 13$ kDa) by size-exclusion chromatography. Crystals of uncomplexed native NolR and selenomethionine (SeMet)-substituted NolR in complex with a 22-bp oligonucleotide duplex corresponding to the consensus NolR DNA binding motif (oligo AT) were obtained and optimized for data collection (Table S1). The 3D structure of SeMet-substituted NolR bound to oligo AT was determined using single-wavelength anomalous dispersion (SAD) phasing. The resulting model was used to solve the structure of unliganded NolR by molecular replacement.

The overall structure of NolR reveals that the protein is a winged helix-turn-helix transcription factor (25) (Fig. 1 *A* and *B*). Two α-helices (α1 and α5) of each monomer form the coiled-coil dimerization interface of NolR. A triangular set of α-helices (α2–α4) positions α3 (residues 45–52) and α4 (residues 55–69) as the helix-turn-helix motif for interaction with the DNA major groove, and the "wing," a two-stranded antiparallel β-sheet (β1a and β1b), extends into the minor groove (Fig. 1*B*). For cocrystallization, the conserved NolR operator site from *R. meliloti* was used (Fig. 1*C*) (17). The operator contains two conserved motifs with variable positions that can be either A or T. Comparison of the free and bound forms of NolR indicates that the structure changes little upon DNA binding with a 0.433-Å root-mean-square deviation (rmsd) for 196 Cα atoms in the homodimer. The crystal structure reveals that the NolR dimer binds to position residues of α4 from one monomer to contact the first sequence block (Fig. 1 *B* and *C*, purple) on the 5′ strand of the operator and that the same helix of the second monomer interacts with the sequence of the second consensus block (Fig. 1 *B* and *C*, red) on the 3′ strand.

Sequence and structural comparisons identify NolR as a member of the ArsR/SmtB family of transcription factors (26–29). NolR shares 22–40% amino acid sequence identity with BigR, HlyU, CadC, CzrA, NmtR, and SmtB (Fig. S1*A*). Moreover, the secondary structure features forming the helix-turn-helix motifs of these proteins are highly conserved (Fig. S1*A*). A search of the Protein Data Bank using the DALI server (30) identifies BigR, HylU, and CadC as the closest structural relatives of NolR (Fig. S1 *B–D*) with Z-scores of 14.4–14.9 and 2.0–2.7 Å rmsd for 95–97 Cα atoms (26, 31–33). The major differences between these proteins occur in the length and positioning of the N-terminal α-helical region at the dimerization interface (Fig. S1 *B–D*). Residues of the regulatory metal-binding sites found in other ArsR/SmtB family members are missing from NolR (26–29, 31–33).

**Asymmetric Operator Site Recognition by NolR.** The structure of NolR complexed with DNA (Fig. 1 *B* and *C*) provides detailed information on how this homodimeric protein recognizes an asymmetric operator site to regulate expression of nodulation and symbiosis genes. Electron density for oligo AT DNA bound to NolR was well-defined (Fig. S2*A*). NolR binds to the operator, with the α4 helix of each monomer positioned 39 Å apart within major grooves of the DNA duplex (Fig. 1*B*). The molecular surface of NolR along the DNA-binding interface of each monomer is positively charged and provides an interaction surface for

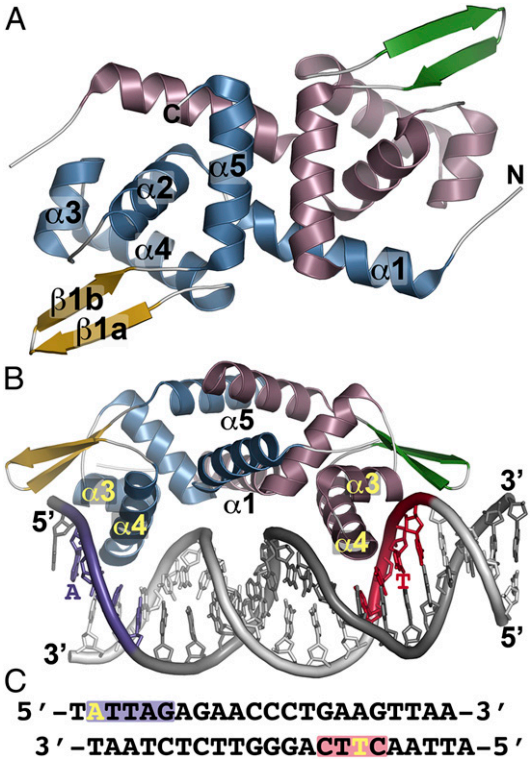


**Fig. 1.** Overall structure of NolR. (*A*) The structure of unliganded NolR is shown as a ribbon diagram. Secondary structure features are labeled in monomer A and are differentially colored in each monomer as follows: blue α-helices and gold β-strands in monomer A and rose α-helices and green β-strands in monomer b. This is a "top" view of the dimeric structure. (*B*) Structure of NolR in complex with the 22-bp oligo AT duplex. Secondary structure features are colored as in *A*. The view is rotated ~90° relative to *A* to present a "side" view of interaction with DNA. α-Helices forming the dimer interface (α1 and α5) and the helix-turn-helix motif (α3–α4) are labeled. Consensus-motif regions of oligo AT that contact NolR are colored purple and red with key nucleotides indicated. (*C*) Sequence of oligo AT. The purple and red boxes correspond to the regions of the consensus motif highlighted in *B*. The yellow A and T indicate nucleotides that are variable in the target DNA sequences of NolR.

phosphate groups of the DNA whereas the opposite side of NolR is largely negative in charge (Fig. S2*B*). Analysis of the operator DNA geometry using 3D-DART (34) shows that the duplex bends 16.8° from an ideal B-form upon interaction with NolR (Fig. S2*C*). Each monomer of the NolR dimer interacts with a sequentially distinct half-site on the operator primarily through residues on α4 (Fig. 2).

Within the first block of the consensus sequence (Fig. 2*A*, purple), two clusters of residues from chain A form extensive interactions with the phosphate backbone of each DNA chain (Fig. 2 *A* and *B*). Gln79, Ile81, and Tyr83 hydrogen bond with T1 and A2 of the 5′ strand. On the 3′ strand, Asn28, Arg31, and His62 contact the phosphates of T17' and C18'. Gln56, Ser57, Ser60, and Gln61 provide hydrogen-bond interactions with T1, A2, T3, and T4 of the 5′ strand. Ser57 and Gln61 hydrogen bond with T19' and A20' of the complementary strand. At the second half site, (Fig. 2*A*, red) a similar set of residues from monomer B provides additional protein–DNA contacts (Fig. 2 *A* and *C*). Gly46, Arg67, Ile81, and Tyr83 of chain B interact with the phosphate groups of C6', T7', and T8' on the 3′ strand. In the second site, Ser60 from α4 hydrogen bonds to the phosphate of T7' instead of a nucleotide base. Interactions with the 5′-strand phosphates of C12 and C13 are contributed by Asn28, Lys30, Arg31, and His62. The β-sheet wing of monomer B binds in the minor groove to place Gln79, which is on the loop between the two β-strands, within hydrogen-bond distance of A21 and A22 of the 5′ strand (Fig. 2*A* and Fig. S3). Ser57 of

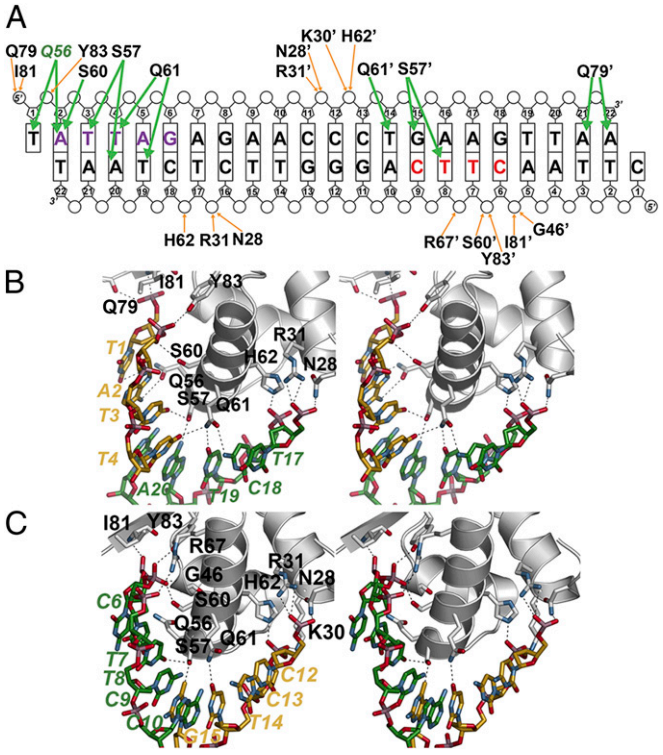


**Fig. 2.** NolR and asymmetric operator binding. (*A*) Schematic showing NolR-oligo AT DNA contacts. The bases are labeled and shown as rectangles, with phosphate and ribose groups drawn as circles and pentagons, respectively. Residues from chain B of the homodimer are noted with an apostrophe after the amino acid number. Orange arrows indicate backbone contacts, and green arrows show base-specific interactions. The two halves of the NolR consensus target sequence that interact with NolR are highlighted with purple and red color, as in Fig. 1 *B* and *C*. Gln56 is colored green to emphasize its role in consensus-motif recognition. (*B*) Stereoview of protein–DNA interactions in the first half-site. Protein side-chains are from chain A. Nucleotides from 5′ and 3′ strands are colored gold and green, respectively, and are labeled. (*C*) Stereoview of protein–DNA interactions in the second half-site. Protein side-chains are from chain B. Nucleotides from 5′ and 3′ strands are colored gold and green, respectively, and are labeled.

monomer B provides bridging contacts with the nucleotide bases of T8' on the 3′ strand and G15 of the 5′ strand. These interactions are similar to those observed in chain A. The side-chain of Gln61 hydrogen bonds to T14 of the 5′ strand. Interestingly, the side-chain of Gln56 in chain B adopts a conformation that flips the amide group away from T7' of the 3′ strand. In contrast, the side-chain of Gln56 in the first site is oriented toward the A2 adenine ring.

**Movement of Gln56 in Recognition of Variable DNA Sequence.** The shift in position of Gln56 in each half-site of the oligo AT DNA, either to interact with A2 in the first site or away from T7' in the second site, suggested that movement of this residue may play a role in the previously observed recognition by NolR of variable operator sites (17–24). To examine the possible role of Gln56 as a conformational switch in DNA binding, NolR was crystallized with oligo AA DNA (Fig. 3*A*). The oligonucleotides used to form this duplex maintain the sequence of the first half-site but substitute a T for A17 and an A for T7' on the 5′ and 3′ strands, respectively. The 3.0-Å resolution crystal structure of NolR in complex with oligo AA was solved by molecular replacement (Table S1). The overall structures of NolR with either DNA bound are nearly identical with a 0.2-Å rmsd for 196 Cα atoms in the homodimer. Comparison of the orientation of Gln56 in the DNA binding sites of each structure show that the interaction with A2 in the first half-site was identical (Fig. 3*B* and Fig. S4). In contrast, the side-chain of Gln56 in the second site, which was oriented away from T7' in the NolR•oligo AT crystal structure, flips to position the amide group for interaction with A7' in the NolR•oligo AA complex (Fig. 3 *B* and *C*). Movement of Gln56 to accommodate the variable A/T positions in each half-site of the NolR consensus DNA binding sequence provides a mechanism for recognition of diverse operator sites of genes involved in nodulation and symbiosis of rhizobia and plants (17).

**Analysis of DNA Binding by NolR Using Isothermal Titration Calorimetry.** The X-ray crystal structures of NolR bound to oligo AT (Fig. 2) and oligo AA (Fig. 3) provide molecular insight on how changes in the position of Gln56 allows for interaction with operator sites with varied sequence at key positions. To examine NolR binding to varied target sites, isothermal titration calorimetry (ITC) was used to characterize interaction with oligos AT and AA. Binding of each DNA duplex to NolR was observed by ITC (Fig. S5); however, the energetic contributions to protein–DNA interaction varied for each operator sequence (Table 1). Analysis of DNA binding to NolR indicated binding of one dimer per operator site, as observed crystallographically, with comparable affinity ($K_d \sim 0.4$ μM). Interestingly, the thermodynamics of interaction were distinct. NolR binding to oligo AT displayed a greater entropic contribution to protein–DNA interaction whereas association with oligo AA was dominated by enthalpic energy. These data indicate that differences in operator site sequence do not alter binding affinity but do change the energetics driving NolR–DNA interaction.

The structures of NolR also implicate key residues as important for DNA binding. To examine the contribution of Arg31, Gln56, Ser57, Ser60, and Gln61, a series of site-directed mutants

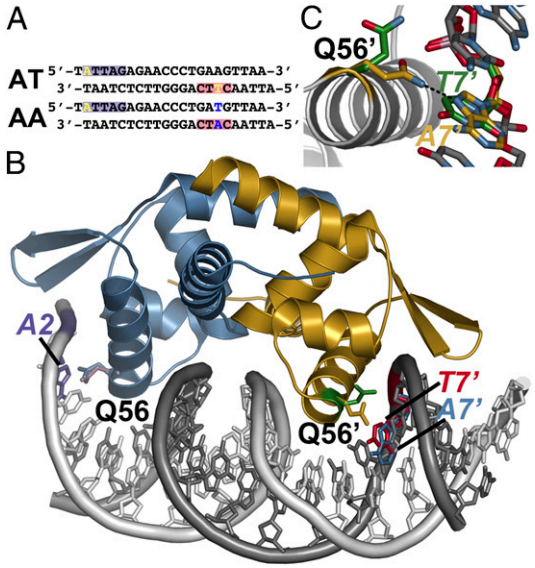


**Fig. 3.** Conformational switching of Gln56 for recognition of variable DNA target sites by NolR. (*A*) Sequences of oligos AT and AA. The purple and red boxes correspond to the regions of the consensus motif highlighted in Fig. 1 *B* and *C*. The variable position bases are highlighted yellow in oligo AT. Changes at these positions in oligo AA are highlighted in blue. (*B*) Structure of NolR in complex with the 22-bp oligo AA duplex. Chains A and B of NolR are colored blue and gold, respectively. The 5′ and 3′ strands of oligo AA are colored white and gray, respectively. The orientation of the Gln56 side-chain in each monomer of NolR complexed with either oligo AT (purple and green sticks) or oligo AA (blue and gold sticks) is shown. The positions of A2 (purple) from oligo AA, T7′ (red) from oligo AT, and A7′ (blue) from oligo AA are shown. (*C*) Close-up view of Gln56 movement in the second consensus half-site of NolR. The position of Gln56 of chain B and the variable base is shown. The structures observed with NolR complexed with either oligo AT (green) and oligo AA (gold) are shown.

**Table 1. Thermodynamic parameters of DNA binding to wild-type NolR**

| DNA | *n* | $K_d$, μM | ΔG, kcal·mol$^{-1}$ | ΔH, kcal·mol$^{-1}$ | −TΔS, kcal·mol$^{-1}$ |
|---|---|---|---|---|---|
| AT | 1.02 ± 0.01 | 0.43 ± 0.06 | −8.69 ± 1.11 | −3.32 ± 0.04 | −5.37 |
| AA | 0.98 ± 0.01 | 0.36 ± 0.03 | −8.78 ± 0.69 | −6.67 ± 0.05 | −2.11 |

Titrations were performed using 22-bp DNA duplexes.

were generated, expressed, and purified. The Q56A mutation was designed to remove the mobile side-chain from NolR for examination of the energetics of protein–DNA interaction by ITC. Arg31, which provides a charge–charge interaction with the DNA backbone, is invariant across members of the ArsR/SmtB family (Fig. S1*A*). Ser57, which in NolR is positioned for bridging interactions between each DNA strand, is also highly conserved (Fig. S1*A*). Ser60 provides either a base contact (site 1) or interacts with the phosphate backbone (site 2) and is invariant in the ArsR/SmtB family (Fig. S1*A*). Gln61 of NolR is either a glutamine (NolR, BigR, HlyU) or a histidine (CadC, CzrA, NmtR, SmtB) in the ArsR/SmtB family (Fig. S1*A*).

The NolR R31A, S57A, S60A, and Q61A mutant proteins were soluble and migrated as dimeric species, as observed for wild-type protein; however, each protein displayed a loss of DNA binding based on the lack of heat signal observed by ITC. These results indicate that these residues in NolR provide critical interactions for formation of protein–DNA complexes. In contrast, binding of oligos AT and AA to the NolR Q56A mutant was observed (Fig. S6). Substitution of Gln56 with an alanine results in less than twofold changes in the $K_d$ values for each DNA duplex compared with wild-type NolR (Table 2). Removal of the amide side-chain in the G56A mutant results in comparable enthalpic and entropic contributions to DNA binding of each duplex. This result is consistent with the movement of Gln56 accounting for the energetic differences for variable DNA operator sequences observed with wild-type NolR.

## Discussion

Although the symbiosis between rhizobial microbes and host plants that leads to nitrogen-fixing nodules is an ecologically and agriculturally important process (1–3), the molecular basis underlying the transcriptional regulation of nodulation and symbiosis remains incompletely understood. Nodulation requires induction of *nod* gene expression; however, efficient symbiosis with host plants occurs only when these genes are expressed in an appropriate quantitative, spatial, and temporal pattern and involves both positive and negative regulation (7, 8, 35). Mutations that alter either positive or negative regulation of *nod* genes result in aberrant and delayed nodulation phenotypes.

As part of the transcriptional regulation machinery of nodulation and symbiosis in various Rhizobium, NolR was originally identified as a putative helix-turn-helix family member that bound a nonpalindromic consensus motif in the core *nodABC* gene cluster (17). Later studies broadly implicate NolR as a global regulator of multiple symbiosis-related genes. The crystal structure of NolR is consistent with DNA footprinting studies that mapped the operator site as containing the (A/T)TTAG-$N_9$-A(T/A) consensus motif (17); however, the NolR structures obtained in complex with two different DNA duplexes (Figs. 1–3) more accurately define the interaction sequences in each asymmetric half site as ATTAG on the 5′ strand and CTTC on the 3′ strand. A remarkable feature of NolR is that the homodimer binds to an asymmetric operator site with variable sequences at two positions, one in each half site, which allows versatility in recognition of NolR operator sites of multiple target genes (19, 20, 22–24).

The 3D structures of NolR in complex with oligo AT (Fig. 2) and oligo AA (Fig. 3) provide new insight on protein–DNA interactions that regulate nodulation and symbiosis gene expression. Binding contacts with the phosphate backbone are largely contributed by residues on α2, α3, and the β-wing. These contacts likely drive nonspecific association with the operator. Base-specific interactions come from Gln56, Ser57, Ser60, and Gln61 on α4 and Gln79 on the β-wing (Fig. 2 and Fig. S3). The hydrogen-bond network between residues of α4 in each half-site and the DNA duplex shows conserved interactions for operator binding. Ser57 anchors protein interaction by contacting T3 in the first half-site and T8’ in the second half-site (Fig. 2). Ser57 also provides hydrogen bonds with A20’ in site one and G15 in site two. Gln61 supports the bridging interactions of Ser57 at both sides of the operator. The amide of Gln61 hydrogen bonds with T19’ and T14 in the first and second half-sites, respectively. ITC analysis of the NolR S57A and Q61A mutants, which showed a lack of interaction, confirms these residues role in DNA binding. Comparison of NolR operators in *nodD1*, *nodZ*, *nolR*, *nodABC*, and *ttsI* with oligos AT and AA reveals that this central set of interactions would likely be conserved across recognition sites (Table S2). In particular, the high conservation of the first half-site suggests that this site is critical for driving NolR–DNA interaction and that nucleotide variation in the second half-site leads to differences in repression by NolR; however, further studies are required to examine how sequence variations potentially modify NolR binding interactions and/or alter target-gene expression.

In addition to these critical contacts with the core consensus motifs, movement of Gln56 allows for accommodation of variable sequences at A2 and T7’ (Fig. 3). The structure of NolR with oligo AT showed that the Gln56 side-chain hydrogen bonds with A2 in the first half-site (Fig. S4) but is flipped away from T7’ in the second site (Fig. 3). Substitution of T7’ with A7’ in oligo AA provides an additional hydrogen bond interaction that orients Gln56 toward the base (Fig. 3). Although NolR binds both operator sequences with comparable affinity, the energetics of protein–DNA interaction differs between the two oligonucleotides (Fig. S5 and Table 1). Interaction with oligo AT displays a larger contribution from entropy compared with oligo AA, but introduction of an additional hydrogen-bond interaction with A7’ in oligo AA enhances the enthalpic component of binding. Mutation of Gln56 to an alanine did not significantly alter the affinity of NolR for either oligo AT or oligo AA (Fig. S6 and Table 2) but led to a decreased enthalpic contribution to binding of oligo AA. Movement of Gln56 leads to compensatory energetic effects that maintain the binding affinity of NolR for varied operator sites.

The X-ray crystal structures of NolR in complex with DNA also provide a first view of protein–DNA interaction in the ArsR/SmtB transcription factor family (Fig. S1). Unlike other ArsR/SmtB transcription factors (26–29, 31–33), no global structural

**Table 2. Thermodynamic parameters of DNA binding to the NolR Q56A mutant**

| DNA | *n* | $K_d$, μM | ΔG, kcal·mol$^{-1}$ | ΔH, kcal·mol$^{-1}$ | −TΔS, kcal·mol$^{-1}$ |
|---|---|---|---|---|---|
| AT | 1.00 ± 0.01 | 0.26 ± 0.01 | −8.99 ± 0.46 | −3.33 ± 0.01 | −5.66 |
| AA | 1.03 ± 0.01 | 0.82 ± 0.11 | −8.30 ± 1.09 | −3.77 ± 0.07 | −4.53 |

Titrations were performed using 22-bp DNA duplexes.

changes in the unliganded and DNA bound forms of NolR were observed, which likely results from a lack of metal-binding sites in the dimer interface of NolR. In the ArsR/SmtB proteins, metal binding triggers conformational changes that result in derepression of gene expression by driving repositioning of the helix-turn-helix DNA interaction motif (26–29, 31–33). This movement results in a switch from a "closed" DNA-binding structure to an "open" low-affinity conformation of the homodimer. To date, there is no evidence for other ligand interactions with NolR that trigger conformational changes that alter binding of the operator site.

Previous structural studies of ArsR/SmtB proteins largely focused on the role of metal-dependent conformational changes and not protein–DNA interaction, with one exception. NMR and mutagenesis probed residues in CzrA from *Staphylococcus aureus* for roles in binding to the palindromic 28-bp *czr* operator (28). This work suggests that CzrA interaction with target DNA increases protein motion in the allosteric sites and showed essential roles for residues analogous to Gln56, Ser57, Ser60, and Gln61 in DNA binding by CzrA, but did not reveal the details of protein–DNA contacts required for phosphate backbone interactions and/or base specificity.

At the molecular level, interaction of NolR with operator sites of genes related to nodulation and symbiosis suggests at least two models for how the repressor functions. The first model of NolR regulation of gene expression is that of binding to an operator within the transcription initiation site of a target gene (Fig. 4*A*). The NolR binding sites for *nodD1*, *nodZ*, and *nolR* (Table S2) are all examples of this arrangement. NodD1 is the major transcription factor for nodulation and NodZ is an α1,6-fucosyltransferase for Nod-factor assembly (7–9, 14, 35–40). In each gene, a NolR operator is positioned 15–60 bp upstream of the coding region (23). Likewise, the presence of a NolR binding site 35 bp in front of the *nolR* gene suggests that levels of NolR modulate its own expression. Increased expression of NolR in free-living Rhizobia down-regulates expression of *nodD1* and other nodulation/symbiosis genes, such as *nodZ* (22). In this arrangement, NolR binding to the operator would compete with RNA polymerase in the promoter site (Fig. 4*A*).

A second model of action involves the binding of NolR to its operator site to alter how NodD either recognizes the *nod* box or how it interacts with RNA polymerase at the transcription initiation site (Fig. 4*B*). NolR operator sites are present in the regulatory regions of the *nodABC* operon and the *ttsI* gene (Table S2) (23). The role of NodD as a key regulator of *nodABC* gene expression is well explored at the genetic level (39). Binding of NodD to the *nod* box sequence 218 bp upstream of *nodABC* activates gene expression. Similarly, NodD binding to the *nod* box 295 bp upstream of the *ttsI* gene controls expression of a protein that in turn activates expression of genes associated with symbiosis (i.e., *nopX*, *nopA, rhcJ* and *rhcQ*) (41–43). Bending of DNA following NodD binding to the *nod* box has been suggested as important for activating expression of genes downstream (13). Multiple studies suggest that binding of LTTR transcription factors at activator sites alters DNA structure to allow for either direct interaction with RNA polymerase or to enhance RNA polymerase promoter escape by DNA bending (10). In the case of the *nodABC* gene cluster, the NolR operator overlaps with the *nod* box (17, 23). Binding of NolR at the operator may either physically compete with NodD at the *nod* box or alter bending at the promoter to down-regulate expression of downstream nodulation genes (Fig. 4*B*). The structures of NolR complexed with DNA suggest that the β-wing and α3 would likely sterically interfere with NodD binding at the *nod* box. It is possible that bending of the operator by NolR (Fig. S2*C*) either rigidifies the DNA to alter interaction with the promoter or changes the structure of the *nod* box to modulate NodD interaction. Further studies are needed to fully understand how these opposing transcription factors modulate gene expression.

In conclusion, the structural and thermodynamic studies presented here provide, to our knowledge, the first insights on the molecular foundation for the regulation of gene expression in nodulation and symbiosis and suggest how the global regulatory protein NolR recognizes variable asymmetric operator sites in the promoter regions of diverse rhizobial genes.

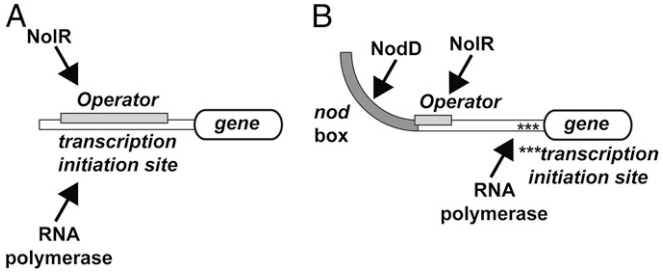


**Fig. 4.** Models of NolR regulation of nodulation and symbiosis gene expression. (*A*) In promoters with overlapping transcription initiation and operator sites, NolR binding prevents RNA polymerase interaction and gene expression. (*B*) In promoter regions containing upstream *nod* box sequences for binding of the transcriptional activator NodD, binding of NolR to the operator site may either alter association of NodD to the *nod* box or alter DNA bending that results from NodD binding to prevent activation of gene expression.

## Methods

**Protein Expression, Purification, and Mutagenesis.** The coding region of *nolR* was PCR-amplified from *S. fredi*i USDA191 genomic DNA using oligonucleotide primers that included NdeI and XhoI restriction sites, respectively, to facilitate cloning. The PCR product was digested with NdeI and XhoI and ligated into pET-28a. The resulting vector was then transformed into *E. coli* BL21 (DE3) cells. Transformed *E. coli* cells were grown at 37 °C in Terrific broth containing 50 $\mu g \cdot mL^{-1}$ kanamycin until $A_{600nm} \sim 0.6–0.9$. After induction with 0.1 mM isopropyl 1-thio-β-D-galactopyranoside, the cultures were grown at 20 °C overnight. Following centrifugation, the cell pellet was suspended in 50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, 1 mM β-mercaptoethanol (βME), 10% (vol/vol) glycerol and 1% Tween 20. Sonication was used for cell lysis. After centrifugation, the supernatant was passed over an $Ni^{2+}$-nitriloacetic acid column. The column was washed with buffer minus Tween 20, and the bound His-tagged protein was eluted using 250 mM imidazole in wash buffer. The N-terminal His-tag was removed by thrombin digestion in dialysis against wash buffer. A mixed $Ni^{2+}$-nitriloacetic acid/benzamidine Sepharose column was used to remove undigested protein and thrombin. Size-exclusion chromatography was performed using a Superdex-75 26/60 FPLC column equilibrated in 150 mM Tris (pH 8.5), 100 mM NaCl, and 5 mM βME. Fractions corresponding to the protein peak were pooled and concentrated to 10 $mg \cdot mL^{-1}$ by centrifugal filtration. Protein concentration was determined using the Bradford method, with BSA as a standard. SeMet-substituted NolR protein was produced by inhibition of the *E. coli* methionine biosynthesis pathway (44) and was purified as described above. Site-directed mutants of NolR (R31A, Q56A, S57A, S60A, and Q61A) were generated using the QuikChange PCR method and were expressed and purified using the same methods used with wild-type NolR.

**Protein Crystallography.** All NolR protein crystals were grown by the hanging drop vapor diffusion method at 4 °C. For crystallization of NolR in the presence of oligo AT, two 22-bp-long oligonucleotides (5′-dTATTAGAGAACCCTGAAGTTAA-3′ and 5′-dATTAACTTCAGGGTTCTCTAAT-3′) were suspended in purification buffer and annealed. Crystals of SeMet-substituted NolR (10.0 $mg \cdot mL^{-1}$) in complex with oligo AT DNA (1 mM) grew in drops containing a 1:1 mixture of protein and crystallization buffer [20% PEG-3350, 0.1 M sodium citrate/citric acid, and 0.2 M sodium citrate (pH 4.0)]. Crystals of native NolR (10 $mg \cdot mL^{-1}$) were formed in drops of a 1:1 mixture of protein and crystallization buffer [1.6 M sodium phosphate monobasic/0.4 M potassium phosphate dibasic, 0.1 M sodium phosphate dibasic/citric acid (pH 4.2)]. Crystals of native NolR (10 $mg \cdot mL^{-1}$) in complex with 1 mM oligo AA DNA (1 mM) (5′-dTATTAGAGAACCCTGATGTTAA-3′ and 5′-dATTAACATCAGGGTTCTCTAAT-3′) were obtained in conditions similar to that of the NolR oligo AT complex. All crystals were stabilized in crystallization solution with 30% glycerol before flash freezing in liquid nitrogen for data collection at 100 K. All X-ray diffraction data (wavelengh = 0.979 Å) were collected at

beamline 19-ID of the Argonne National Laboratory Advanced Photon Source. HKL3000 (45) was used to index, integrate, and scale diffraction data. The structure of SeMet-substituted NolR in complex with oligo AT DNA was determined by SAD phasing. SHELX (46) was used to determine SeMet positions and to estimate initial phases from the peak wavelength dataset. Refinement of SeMet positions and parameters was performed with MLPHARE (47). Solvent flattening using density modification implemented with ARP/wARP (48) was used to build an initial model. Subsequent iterative rounds of manual model building and refinement, which included translation-libration-screen parameter refinement, used COOT (49) and PHENIX (50), respectively. The structures of unliganded NolR and NolR complexed with oligo AA DNA were solved by molecular replacement in PHASER (51) using the SeMet-substituted NolR structure as a search model, with model building and refinement performed with COOT and PHENIX. Waters were added to the unliganded NolR model using default parameters in PHENIX. Crystallographic statistics are summarized in Table S1.

**Isothermal Titration Calorimetry.** NolR protein was dialyzed overnight in 150 mM Tris (pH 8.5), 100 mM NaCl, 5 mM βME, 5 mM $MgCl_2$, and 5% glycerol at 4 °C. Synthetic double-stranded DNA (oligo AT, AA, and TT) were prepared in the same buffer. ITC experiments were performed using a VP-ITC calorimeter (Microcal) at 4 °C. Data obtained from the titrations were analyzed using a single-site binding model: $Q_i^{tot} = V_0 \bullet M_i^{tot} \bullet ((nK_1x)\Delta H_1)/(1 + K_1x)$, in which $Q_i^{tot}$ is the total heat after the $i^{th}$ injection, $V_0$ is the calorimetric cell volume, $M_i^{tot}$ is the concentration of protein in the cell after the $i^{th}$ injection, $\Delta H$ is the corresponding enthalpy change to NolR•DNA binding, $n$ is the number of nucleotide binding sites, and $K$ is the equilibrium binding constant. Estimates of $K_{obs}$ and $\Delta H$ were obtained by fitting the experimental data using Origin software (Microcal). Values for the change in free energy ($\Delta G$) were calculated using $\Delta G = -RT\ln(K_{obs})$, where R is the gas constant (1.9872 $cal \cdot K^{-1} \cdot mol^{-1}$) and T is absolute temperature. Changes in entropy ($\Delta S$) were calculated using $\Delta G = \Delta H - T\Delta S$. $K_d$ was calculated as $1/K_{obs}$.

**ACKNOWLEDGMENTS.** Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center of the Advanced Photon Source, a national user facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research (DE-AC02-06CH11357).

1. Desbrosses GJ, Stougaard J (2011) Root nodulation: A paradigm for how plant-microbe symbiosis influences host developmental pathways. *Cell Host Microbe* 10(4):348–358.
2. Kondorosi E, Mergaert P, Kereszt A (2013) A paradigm for endosymbiotic life: Cell differentiation of Rhizobium bacteria provoked by host plant factors. *Annu Rev Microbiol* 67:611–628.
3. Masson-Boivin C, Giraud E, Perret X, Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: How many rhizobium recipes? *Trends Microbiol* 17(10):458–466.
4. Horvath B, et al. (1986) Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. *Cell* 46(3):335–343.
5. Long SR (1996) Rhizobium symbiosis: Nod factors in perspective. *Plant Cell* 8(10): 1885–1898.
6. Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC (2007) How rhizobial symbionts invade plants: The *Sinorhizobium-Medicago* model. *Nat Rev Microbiol* 5(8): 619–633.
7. Kondorosi E, et al. (1989) Positive and negative control of nod gene expression in *Rhizobium meliloti* is required for optimal nodulation. *EMBO J* 8(5):1331–1340.
8. Loh J, Stacey G (2003) Nodulation gene regulation in *Bradyrhizobium japonicum*: A unique integration of global regulatory circuits. *Appl Environ Microbiol* 69(1):10–17.
9. Göttfert M, et al. (1986) At least two *nodD* genes are necessary for efficient nodulation of alfalfa by *Rhizobium meliloti*. *J Mol Biol* 191(3):411–420.
10. Maddocks SE, Oyston PC (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* 154(Pt 12):3609–3623.
11. Fisher RF, Brierley HL, Mulligan JT, Long SR (1987) Transcription of *Rhizobium meliloti* nodulation genes: Identification of a *nodD* transcription initiation site in vitro and in vivo. *J Biol Chem* 262(14):6849–6855.
12. Fisher RF, Egelhoff TT, Mulligan JT, Long SR (1988) Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. *Genes Dev* 2(3):282–293.
13. Fisher RF, Long SR (1993) Interactions of NodD at the nod Box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA. *J Mol Biol* 233(3):336–348.
14. Winsor BAT (1989) A nod at differentiation: The *nodD* gene product and initiation of Rhizobium nodulation. *Trends Genet* 5(7):199–201.
15. Peck MC, Fisher RF, Long SR (2006) Diverse flavonoids stimulate NodD1 binding to nod gene promoters in *Sinorhizobium meliloti*. *J Bacteriol* 188(15):5417–5427.
16. Peck MC, Fisher RF, Bliss R, Long SR (2013) Isolation and characterization of mutant *Sinorhizobium meliloti* NodD1 proteins with altered responses to luteolin. *J Bacteriol* 195(16):3714–3723.
17. Kondorosi E, et al. (1991) Identification of NolR, a negative transacting factor controlling the nod regulon in *Rhizobium meliloti*. *J Mol Biol* 222(4):885–896.
18. Li F, Hou B, Hong G (2008) Symbiotic plasmid is required for NolR to fully repress nodulation genes in *Rhizobium leguminosarum* A34. *Acta Biochim Biophys Sin (Shanghai)* 40(10):901–907.
19. Chen H, et al. (2000) Identification of *nolR*-regulated proteins in *Sinorhizobium meliloti* using proteome analysis. *Electrophoresis* 21(17):3823–3832.
20. Chen H, Gao K, Kondorosi E, Kondorosi A, Rolfe BG (2005) Functional genomic analysis of global regulator NolR in *Sinorhizobium meliloti*. *Mol Plant Microbe Interact* 18(12):1340–1352.
21. Cren M, Kondorosi A, Kondorosi E (1994) An insertional point mutation inactivates NolR repressor in *Rhizobium meliloti* 1021. *J Bacteriol* 176(2):518–519.
22. Cren M, Kondorosi A, Kondorosi E (1995) NolR controls expression of the *Rhizobium meliloti* nodulation genes involved in the core Nod factor synthesis. *Mol Microbiol* 15(4):733–747.
23. Vinardell JM, et al. (2004) NolR regulates diverse symbiotic signals of *Sinorhizobium fredii* HH103. *Mol Plant Microbe Interact* 17(6):676–685.
24. López-Baena FJ, et al. (2008) Regulation and symbiotic significance of nodulation outer proteins secretion in *Sinorhizobium fredii* HH103. *Microbiology* 154(Pt 6):1825–1836.
25. Brennan RG (1993) The winged-helix DNA-binding motif: Another helix-turn-helix takeoff. *Cell* 74(5):773–776.
26. Guimarães BG, et al. (2011) Plant pathogenic bacteria utilize biofilm growth-associated repressor (BigR), a novel winged-helix redox switch, to control hydrogen sulfide detoxification under hypoxia. *J Biol Chem* 286(29):26148–26157.
27. Eicken C, et al. (2003) A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. *J Mol Biol* 333(4):683–695.
28. Arunkumar AI, Campanello GC, Giedroc DP (2009) Solution structure of a paradigm ArsR family zinc sensor in the DNA-bound state. *Proc Natl Acad Sci USA* 106(43): 18177–18182.
29. Osman D, Cavet JS (2010) Bacterial metal-sensing proteins exemplified by ArsR-SmtB family repressors. *Nat Prod Rep* 27(5):668–680.
30. Holm L, Rosenström P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545–W549.
31. Ye J, Kandegedara A, Martin P, Rosen BP (2005) Crystal structure of the *Staphylococcus aureus* pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. *J Bacteriol* 187(12):4214–4221.
32. Kandegedara A, Thiyagarajan S, Kondapalli KC, Stemmler TL, Rosen BP (2009) Role of bound Zn(II) in the CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. *J Biol Chem* 284(22): 14958–14965.
33. Nishi K, et al. (2010) Crystal structure of the transcriptional activator HlyU from *Vibrio vulnificus* CMCP6. *FEBS Lett* 584(6):1097–1102.
34. van Dijk M, Bonvin AMJJ (2009) 3D-DART: A DNA structure modelling server. *Nucleic Acids Res* 37(Web Server issue):W235-9.
35. Knight CD, Rossen L, Robertson JG, Wells B, Downie JA (1986) Nodulation inhibition by *Rhizobium leguminosarum* multicopy *nodABC* genes and analysis of early stages of plant infection. *J Bacteriol* 166(2):552–558.
36. Horvath B, Bachem CW, Schell J, Kondorosi A (1987) Host-specific regulation of nodulation genes in Rhizobium is mediated by a plant-signal, interacting with the *nodD* gene product. *EMBO J* 6(4):841–848.
37. Honma MA, Ausubel FM (1987) *Rhizobium meliloti* has three functional copies of the *nodD* symbiotic regulatory gene. *Proc Natl Acad Sci USA* 84(23):8558–8562.
38. Mulligan JT, Long SR (1989) A family of activator genes regulates expression of *Rhizobium meliloti* nodulation genes. *Genetics* 122(1):7–18.
39. Roche P, et al. (1996) The common *nodABC* genes of *Rhizobium meliloti* are host-range determinants. *Proc Natl Acad Sci USA* 93(26):15305–15310.
40. Stacey G, et al. (1994) *nodZ*, a unique host-specific nodulation gene, is involved in the fucosylation of the lipooligosaccharide nodulation signal of *Bradyrhizobium japonicum*. *J Bacteriol* 176(3):620–633.
41. Viprey V, Del Greco A, Golinowski W, Broughton WJ, Perret X (1998) Symbiotic implications of type III protein secretion machinery in Rhizobium. *Mol Microbiol* 28(6):1381–1389.
42. Krishnan HB, et al. (2003) Extracellular proteins involved in soybean cultivar-specific nodulation are associated with pilus-like surface appendages and exported by a type III protein secretion system in Sinorhizobium fredii USDA257. *Mol Plant Microbe Interact* 16(7):617–625.
43. Deakin WJ, Marie C, Saad MM, Krishnan HB, Broughton WJ (2005) NopA is associated with cell surface appendages produced by the type III secretion system of *Rhizobium sp*. strain NGR234. *Mol Plant Microbe Interact* 18(5):499–507.
44. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* 229(1):105–124.
45. Otwinowski Z, Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
46. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64(Pt 1):112–122.
47. Terwilliger TC (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56(Pt 8):965–972.
48. Morris RJ, Perrakis A, Lamzin VS (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* 374:229–244.
49. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
50. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
51. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4):658–674.